README.txt
****
README README README README README README README
****

I. Introduction:

THIS SOURCE CODE IS NOT TO BE DUPLICATED WITHOUT PERMISSION

This Disk contains the source code (written in the C/C++ programming language)
for US Patent Application of Pang et al. for "A COMPUTATIONAL METHOD FOR PREDICTING
INTRAMOLECULAR AND INTERMOLECULAR BIOPOLYMER INTERACTIONS"

By Phillip S. Pang, Eckhard Jankowsky and Anna Marie Pyle
For more information contact: phillip.pang@stanfordalumni.org

Filing By: Baker Botts LLP, 30 Rockefeller Plaza, New York, NY, 10112.

II. Files:

The Microsoft Visual C++ 6.0 Programming environment was used:

The workspace file is: Shevek.dsw
The project file is: Shevek.dsp

Source files include:

        ApplyThresh.c
        ChainElim.c
        definitions.h
        FASTEXP2.C
        FASTEXP2.H
        FASTEXP2.INC
        GETTIMER.C
        MainExactPScore.c
        MainMisalign.c
        MainScreening.c
        MainShevek.c
        MisalignNRPang.c
        nrutilp.h
        NumRecPang.c
        NumRecUtilities.c
        PRCERR.C
        RandNumGen.c
        RCONT2p.H
        Rcount2Pang.C
        shevek.h
        SUPPORT.C
        SUPPORT.H

    Text files include:

        degenmsg
        eliminatemsg
        misalignmsg
        predictmsg
        scoremsg
        threshmsg
        titlepage

III. General Program Structure:

        FUNCTION   PURPOSE   FOUND IN FILE

        main( )   [driving function]   MainShevek.c
          |
          |
          -------- openfile( )   [reads alignment]   MainShevek.c
          |
          -------- score_manager( )   [scoring]   MainShevek.c
          |
          -------- screener( )   [screening]   MainScreening.c
          |
          -------- misalign_identifier( ) [misalignment]   MainMisalign.c

```
# Microsoft Developer Studio Project File - Name="shev" - Package Owner=<4>
# Microsoft Developer Studio Generated Build File, Format Version 6.00
# ** DO NOT EDIT **

# TARGTYPE "Win32 (x86) Console Application" 0x0103

CFG=shev - Win32 Debug
!MESSAGE This is not a valid makefile. To build this project using NMAKE,
!MESSAGE use the Export Makefile command and run
!MESSAGE
!MESSAGE NMAKE /f "Shevek.mak".
!MESSAGE
!MESSAGE You can specify a configuration when running NMAKE
!MESSAGE by defining the macro CFG on the command line. For example:
!MESSAGE
!MESSAGE NMAKE /f "Shevek.mak" CFG="shev - Win32 Debug"
!MESSAGE
!MESSAGE Possible choices for configuration are:
!MESSAGE
!MESSAGE "shev - Win32 Release" (based on "Win32 (x86) Console Application")
!MESSAGE "shev - Win32 Debug" (based on "Win32 (x86) Console Application")
!MESSAGE

# Begin Project
# PROP AllowPerConfigDependencies 0
# PROP Scc_ProjName ""
# PROP Scc_LocalPath ""
CPP=xicl6.exe
RSC=rc.exe

!IF  "$(CFG)" == "shev - Win32 Release"

# PROP BASE Use_MFC 0
# PROP BASE Use_Debug_Libraries 0
# PROP BASE Output_Dir "Release"
# PROP BASE Intermediate_Dir "Release"
# PROP BASE Target_Dir ""
# PROP Use_MFC 0
# PROP Use_Debug_Libraries 0
# PROP Output_Dir "Release"
# PROP Intermediate_Dir "Release"
# PROP Target_Dir ""
# ADD BASE CPP /nologo /W3 /GX /O2 /D "WIN32" /D "NDEBUG" /D "_CONSOLE" /D "_MBCS" /YX /FD /c
# ADD CPP /nologo /W3 /GX /O2 /D "WIN32" /D "NDEBUG" /D "_CONSOLE" /D "_MBCS" /YX /FD /c
# ADD BASE RSC /l 0x409 /d "NDEBUG"
# ADD RSC /l 0x409 /d "NDEBUG"
BSC32=bscmake.exe
# ADD BASE BSC32 /nologo
# ADD BSC32 /nologo
LINK32=xilink6.exe
# ADD BASE LINK32 kernel32.lib user32.lib gdi32.lib winspool.lib comdlg32.lib advapi32.lib
shell32.lib ole32.lib oleaut32.lib uuid.lib odbc32.lib odbccp32.lib /nologo /subsystem:console
/machine:I386
# ADD LINK32 kernel32.lib user32.lib gdi32.lib winspool.lib comdlg32.lib advapi32.lib shell32.lib
ole32.lib oleaut32.lib uuid.lib odbc32.lib odbccp32.lib /nologo /subsystem:console /machine:I386

!ELSEIF  "$(CFG)" == "shev - Win32 Debug"

# PROP BASE Use_MFC 0
# PROP BASE Use_Debug_Libraries 1
# PROP BASE Output_Dir "Debug"
# PROP BASE Intermediate_Dir "Debug"
# PROP BASE Target_Dir ""
# PROP Use_MFC 0
# PROP Use_Debug_Libraries 1
# PROP Output_Dir "Debug"
# PROP Intermediate_Dir "Debug"
# PROP Target_Dir ""
# ADD BASE CPP /nologo /W3 /Gm /GX /ZI /Od /D "WIN32" /D "_DEBUG" /D "_CONSOLE" /D "_MBCS" /YX /FD
/GZ /c
# ADD CPP /nologo /W3 /Gm /GX /ZI /Od /D "WIN32" /D "_DEBUG" /D "_CONSOLE" /D "_MBCS" /FR /YX /FD /GZ
/c
# ADD BASE RSC /l 0x409 /d "_DEBUG"
# ADD RSC /l 0x409 /d "_DEBUG"
BSC32=bscmake.exe
# ADD BASE BSC32 /nologo
# ADD BSC32 /nologo
LINK32=xilink6.exe
```

```
# ADD BASE LINK32 kernel32.lib user32.lib gdi32.lib winspool.lib comdlg32.lib advapi32.lib
shell32.lib ole32.lib oleaut32.lib uuid.lib odbc32.lib odbccp32.lib /nologo /subsystem:console /debug
/machine:I386 /pdbtype:sept
# ADD LINK32 kernel32.lib user32.lib gdi32.lib winspool.lib comdlg32.lib advapi32.lib shell32.lib
ole32.lib oleaut32.lib uuid.lib odbc32.lib odbccp32.lib /nologo /subsystem:console /debug
/machine:I386 /pdbtype:sept

!ENDIF

# Begin Target

# Name "shev - Win32 Release"
# Name "shev - Win32 Debug"
# Begin Group "Source Files"

# PROP Default_Filter "cpp;c;cxx;rc;def;r;odl;idl;hpj;bat"
# Begin Group "Primary Files"

# PROP Default_Filter ""
# Begin Source File

SOURCE=.\MainShevek.c
# End Source File
# Begin Source File

SOURCE=.\NumRecPang.c
# End Source File
# Begin Source File

SOURCE=.\NumRecUtilities.c
# End Source File
# End Group
# Begin Group "MonteCarloSim"

# PROP Default_Filter ""
# Begin Source File

SOURCE=.\FASTEXP2.C
# End Source File
# Begin Source File

SOURCE=.\FASTEXP2.INC
# End Source File
# Begin Source File

SOURCE=.\GETTIMER.C
# End Source File
# Begin Source File

SOURCE=.\MainExactPScore.c
# End Source File
# Begin Source File

SOURCE=.\PRCERR.C
# End Source File
# Begin Source File

SOURCE=.\RandNumGen.c
# End Source File
# Begin Source File

SOURCE=.\Rcount2Pang.C
# End Source File
# Begin Source File

SOURCE=.\SUPPORT.C
# End Source File
# End Group
# Begin Group "Screening"

# PROP Default_Filter ""
# Begin Source File

SOURCE=.\ApplyThresh.c
# End Source File
# Begin Source File

SOURCE=.\ChainElim.c
```

```
# End Source File
# Begin Source File

SOURCE=.\MainScreening.c
# End Source File
# End Group
# Begin Group "Misalignment"

# PROP Default_Filter ""
# Begin Source File

SOURCE=.\MainMisalign.c
# End Source File
# Begin Source File

SOURCE=.\MisalignNRPang.c
# End Source File
# End Group
# End Group
# Begin Group "Header Files"

# PROP Default_Filter "h;hpp;hxx;hm;inl"
# Begin Group "MonteCarlo Headers"

# PROP Default_Filter ""
# Begin Source File

SOURCE=.\FASTEXP2.H
# End Source File
# Begin Source File

SOURCE=.\RCONT2p.H
# End Source File
# Begin Source File

SOURCE=.\SUPPORT.H
# End Source File
# End Group
# Begin Source File

SOURCE=.\definitions.h
# End Source File
# Begin Source File

SOURCE=.\nrutilp.h
# End Source File
# Begin Source File

SOURCE=.\shevek.h
# End Source File
# End Group
# Begin Group "Resource Files"

# PROP Default_Filter "ico;cur;bmp;dlg;rc2;rct;bin;rgs;gif;jpg;jpeg;jpe"
# End Group
# End Target
# End Project
```

```
Microsoft Developer Studio Workspace File, Format Version 6.00
# WARNING: DO NOT EDIT OR DELETE THIS WORKSPACE FILE!

###############################################################################

Project: "shev"=".\shev\shev.dsp" - Package Owner=<4>

Package=<5>
{{{
}}}

Package=<4>
{{{
}}}

###############################################################################

Global:

Package=<5>
{{{
}}}

Package=<3>
{{{
}}}

###############################################################################
```

```c
/**********************************************************/
/********      Written By Phillip S. Pang      *******/
/**********************************************************/


/**********************************************************/
/******** MD/PhD Candidate, Columbia University *******/
/******** College of Physicians and Surgeons   *******/
/******** Dept. Of Biochemistry and Biophysics  *******/
/**********************************************************/


/**********************************************************/
/********    phillip.pang@stanfordalumni.org   *******/
/**********************************************************/



/**********************************************************/
/********          STATEMENT OF COPYRIGHT        *******/

/*          Copyright 2001 by The Trustees of       */
/*          Columbia University in the City of      */
/*          New York. ALL RIGHTS RESERVED;          */

/**********************************************************/



/*******************************************************************/
/*******************************************************************/
/*******************************************************************/
/*                                                                 */
/*   Certain algorithms found within this file may be derivatives*/
/*   of source code obtained from the book:                        */
/*   "Numerical Recipes in C: The Art of Scientific Computing"     */
/*   published by Cambridge University Press.                      */
/*                                                                 */
/*******************************************************************/
/*******************************************************************/
/*******************************************************************/



/************************/
/************************/

/**    PROGRAM BEGINS    ***/

/************************/
/************************/



#include <string.h>
#include <stdio.h>
#include <stdlib.h>
#include <ctype.h>
#include <math.h>
#include <time.h>

#include "nrutilp.h"
#include "shevek.h"
#include "definitions.h"


double PRECISION;
int prec;

float    NUM_COL;
int      NUM_ROW;
int      OFFSET;

int      mainseq;
int      *dlength;

int      START;
int      STOP;
```

```
/*****************************************************/
/*****************************************************/
/**********   MAIN FUNCTION ******************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/**********   MAIN FUNCTION ******************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/**********   MAIN FUNCTION ******************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/*****************************************************/
/**********   MAIN FUNCTION ******************/
/*****************************************************/
/*****************************************************/


void main ()
{
    void    intro();
    char    *openfile();
    void    score_manager(char *input);
    void    positionrelater(char *input);
    void    screener(int dfilelength, float *VThresh, float *PThresh);
    void    exitprogram();
    void    misalign_identifier(char *input, int offset, int numrows);


    char    *input;

    int     dl;                      /*memory space for file length of alldata.txt*/

    float VT,PT;
    float *VThresh, *PThresh;

    extern int prec;


    VThresh = &VT;
    PThresh = &PT;
    dlength = &dl;


    intro();                              /*INTRODUCTORY TEXT */

    input = openfile();                   /*OPENS THEN READS sequence alignment file*/
                                          /*returns pointer to sequence data*/

    positionrelater(input);               /*outputs correlation between position number and
residue/nucleotide*/

                                          /*number of primary sequence under analysis*/


    score_manager(input);                 /*SCORES sequence data positions*/
                                          /*outputs data files*/
                                          /*STANDARDIZES scores as necessary*/

    screener(dl,VThresh,PThresh);         /*SCREENING of Scores */
                                          /*1) Analyzes Distribution of Standardized scores*
/
                                          /*2) Suggests lower min. thresholds*/
                                          /*3) Uses thresholds for preliminary list of predi
ctions*/
                                          /*4) Eliminates intersecting interactions */
                                          /*5) Outputs predict.txt file of predictions */


    misalign_identifier(input,OFFSET,NUM_ROW);  /*takes prediction list,and generates*/
                                          /*regenerates actual freq tables -- avoid
```

```c
having to keep them*/
                                                     /*generates AR tables*/
                                                     /*outputs misalign.txt file*/


    exitprogram();


    free_cvector(input, 0, 0);        /*frees memory allocated for     */
                                      /*aligned sequences       */


}


/**********************************************/
/*********** INTRO *********************/
/**********************************************/

void    intro()
{
    int numread;
    int numwrite;
    char buf[10];
    FILE *inputfile = NULL;

    /* opens input file */

    inputfile = fopen("titlepage", "r");

    if (inputfile == NULL)  printf("\nTITLE PAGE NOT FOUND\n");

    else {
    /*read FILE and writes to screen until end of file is reached*/

        while(  !feof(inputfile)   )
        {
          numread =  fread(buf,sizeof(char),1,inputfile);
          numwrite = fwrite(buf,sizeof(char),1,stdout);
        }

        fflush(inputfile);
        fclose(inputfile);
    }
}




/**********************************************/
/*********** openfile *********************/
/**********************************************/

char    *openfile()
{

    char    *read_input(FILE *ifile);
    char    *input;
    char    filename[50];


    extern int  mainseq;
    extern double PRECISION;
    extern int prec;

    extern int START;
    extern int STOP;

    FILE * inputfile = NULL;

    int found_file = 0;

    while (found_file == NO) {                          /*Queries for input file*/
        printf("\nENTER alignment file name (include extension):");
        scanf("%s", filename);
        inputfile = fopen(filename, "r");
        if (inputfile == NULL)    {
```

MainShevek.c 3